5-1-2005

Revealing Socially Undesirable Information: A Comparison of Bipolar Adjective Scaling Methods

Jordan Mudd
Western Kentucky University

REVEALING SOCIALLY UNDESIRABLE INFORMATION: A COMPARISON OF BIPOLAR ADJECTIVE SCALING METHODS

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Arts

By
Jordan Wesley Mudd

May 7, 2005

REVEALING SOCIALLY UNDESIRABLE INFORMATION:

A COMPARISON OF BIPOLAR ADJECTIVE SCALING METHODS

Date Recommended April 18, 2005

Director of Thesis

Dean, Graduate Studies and Research Date

Table of Contents

REVEALING SOCIALLY UNDESIRABLE INFORMATION:

A COMPARISON OF BIPOLAR ADJECTIVE SCALING METHODS

Jordan Wesley Mudd May 7, 2005 38 Pages

Directed by: Reagan Brown, Ph.D., Elizabeth Shoenfelt, Ph.D., Anthony Paquin, Ph.D.

Department of Psychology Western Kentucky University

Despite the popularity of personality testing for personnel selection, criticisms have arisen regarding the potential for response distortions by applicants. Researchers have developed many techniques to control for such response distortions, including the use of different response formats. Using both a university sample of introductory psychology students and an applied sample of police officers during a promotional exam, the present study examined two bipolar adjective scaling methods (paired-comparison and semantic differential) to determine scaling effects on test taker reports of socially undesirable information on a self-report personality measure of conscientiousness. Results indicate that no differences exist in socially desirable responding based on the scaling method used to assess personality. In a follow-up study using college students, a semantic differential scale with six points was administered. Again, no differences were found. Internal consistency analyses comparing a 6-point semantic differential scale with a paired comparison scale indicated greater internal consistency for the semantic differential format.

Introduction

Personality tests are commonly used tools for personnel selection. Despite the popularity of such testing, critics argue that such selection procedures have little utility in practice given that applicants can easily falsify responses in an attempt to present a favorable impression of themselves (Dunnette, McCartney, Carlson, & Kirchner, 1962; Rosse, Stecher, Miller, & Leving, 1998). Many techniques designed to control for response distortion have been proposed in an attempt to increase the accuracy of personality measurement of applicants. This paper examines one such technique, an altered item format. To provide a foundation, the following sections will discuss types of response bias and the research that has examined response bias in testing. Additionally, different types of item formats will be discussed in terms of their advantages and disadvantages. Finally, prior studies will be presented that discuss the impact that item format has on response bias.

Response Bias

Personality inventories are always subject to potential response distortions. Response distortions include social desirability, claiming unlikely virtues, denying common faults and unpopular attitudes, exaggerating personal strengths, good impression, self-enhancement, and faking (Ones, Viswesvaran, & Reiss, 1996). For decades, behavioral scientists have identified the variety of forms in which response bias can be manifested, namely deliberate faking, response sets, or inaccurate estimation of one's own personality (Bartlett, Quay, & Wrightsman, 1960). Research has indicated that response distortion during pre-employment personality testing can have a significant effect on which applicants are ultimately hired (Rosse et al., 1998).

Since the early identification of response distortion, a majority of the attention regarding response distortion has focused on socially desirable responding (Edwards, 1957). Socially desirable responding refers to the tendency of individuals to present themselves in a socially acceptable manner based on current social norms and standards (Zerbe & Paulhus, 1987). Various models have been developed to explain the effects of social desirability on organizational behavior. Ganster, Hennessy, and Luthans (1983) described the relations between social desirability and organizational behavior using three models: social desirability as an unmeasured variable that produces spurious correlations between variables, social desirability as a suppressor variable that hides relationships, and social desirability as a moderator variable that creates an interaction effect between two other variables. More recent meta-analytic research (Ones et al., 1996) has examined social desirability and has shown that it does not function as a predictor, mediator, or suppressor variable in relation to job performance. Rather, Ones et al. argued that social desirability is related to real individual differences in conscientiousness and emotional stability and is not as pervasive a problem as many psychologists have claimed. Job applicants who scored high on social desirability also tended to score high on emotional stability. This argument implies that social desirability should be viewed as a dimension of an individual's personality rather than measurement bias. Although these findings appear to minimize the importance of social desirability, researchers still disagree regarding the effects of social desirability on the validity of personality measurement. Some researchers (Nicholson & Hogan, 1990) argue that if social desirability is not a part of personality (and is just a bias), then personality measured without socially desirable responding would be an improvement.

An alternative approach to social desirability research is to examine the intent behind the response distortion. Although many response distortions are conceptualized as intentional and deceptive, a socially desirable response on a self-report inventory does not necessarily indicate deliberate deception on the part of the respondent. Unintentional deception can occur for a number of reasons including lack of insight into one's characteristics, self-deception, or an unwillingness to face up to one's limitations (Anastasi, 1982). As regards the intent behind the response distortion, Paulhus (1984) proposed a two component model of social desirability: self-deception and impression management.

Self-deceptive positivity is a tendency toward self-deceptive overconfidence in one's strengths and abilities (Strong, Greene, & Kordinak, 2002). The intent behind such responses is evidenced in Paulhus and Reid's (1991) description of self-deception, which they defined as a favorably biased but honestly held self-descriptions. Self-deception has also been seen to play a role in other organizational behavior processes. Self-deception can be a positive self-bias that plays a central role in expectancy theory in which the discounting of failure will lead to a high expectancy of success (Zerbe & Paulhus, 1987). Self-deception can impact decision making. When a fast decision is needed, self-deception can be advantageous because the decision maker will save time by discounting any potential failures. Likewise, when a slower, more reasoned approach is required, self-deception can be dysfunctional because the decision maker will likely have low expectancies for success (Zerbe & Paulhus).

The other component of social desirability is impression management, whereby an individual deliberately attempts to present a socially favorable and overly benign

personality description of himself (Strong et al., 2002). Research on impression management has indicated that some forms, such as denying minor faults, declaring a variety of professional interests, claiming conscientiousness, and behavioral constraint, may not necessarily be identified by traditional validity scales (Strong et al.). Barrick and Mount (1996) examined the effects of impression management and self-deception on the predictive validity of conscientiousness and emotional stability. In order to distinguish between the two types of response bias, Paulhus' (1984) *Balanced Inventory of Desirable Responding* was administered. Its two subscales are designed to measure each type of response bias. The results indicated that both personality dimensions were positively related to supervisory ratings of performance. In addition, results indicated that applicants distorted their scores on both personality dimensions using both self-deception and impression management.

Involved in both components of social desirability is the principle of self-enhancement (or self-presentation). Self-enhancement refers to the tendency to describe oneself in overly positive terms (Robins & Paulhus, 2001). Robins and Paulhus suggest that self-enhancement generally involves two explanations. One explanation involves (a) cognitive processes in which the individual focuses on the information available to the self, (b) prior beliefs and expectancies, and (c) processes of attention, encoding and retrieval of self-relevant information. The other explanation involves motivational or affective factors which focus on the motive to maintain and enhance self-esteem, the desire to reduce negative affect and increase positive affect, and self-presentational concerns such as the need for social approval. Self-enhancement can be viewed as entailing both benefits and costs to the individual and organization (Robins & Paulhus).

Self-presentation in individuals has been shown to be best diagnosed by scales of impression management and conscientiousness (Paulhus, Bruce, & Trapnell, 1995).

Early research by Edwards (1957) identified three approaches to control the effects of social desirability in personality inventories. First, social desirability can be controlled by developing inventories in which the statements are neutral or subtle in terms of social desirability. A second approach consists of the use of a scale that measures socially desirable responding. Scores on a social desirability scale can then be used to statistically correct the scores for the response distortion. Finally, statements can be paired on the basis of their social desirability scale values so that they are equal or close in attractiveness in terms of social desirability. More recent research has suggested the control of social desirability using two main modes (Nederhof, 1985). One method consists of detection and measurement of the social desirability bias through the use of social desirability scales or ratings of item desirability. The other approach consists of a variety of methods, all aimed at preventing or reducing social desirability bias, including the use of forced-choice items, use of proxy subjects, self-administration of the questionnaire, and randomized response technique. Nederhoff examined these approaches to controlling social desirability bias and found that no single method excelled over any other methods. Consequently, he recommended a combination of the methods as the best approach to reduce social desirability bias.

Further research investigated some of these methods of social desirability control in applied settings. Butcher (1994) tested airline pilots using the MMPI-2 to determine the degree of defensiveness in their responses as detected by the L and K scales. The L (Lie) scale is a validity scale that detects overly positive test-taker responses. The K

(Correction) scale is a validity scale that detects test taker defensiveness. The results indicated that the airline pilots presented themselves in a defensive manner. However, the usefulness of such scales was questioned for personnel screening given that the scales were originally developed for other purposes, demonstrated low validity, and lacked the adequate range needed to assess such defensiveness in applicants. Another technique proposed to reduce socially desirable responding is the use of altered instructions during administration on an instrument. For example, altered instructions may consist of explaining to examinees the tendency for responses to be shaped by social desirability and how to avoid such biases in their responses. Anastasi (1982) described the use of altered test instructions and how it may motivate examinees to respond in a frank manner. Cigrang and Stall (2000) examined the use of altered instructions as a way to correct for defensiveness on the MMPI-2. After an initial test, those whose responses were deemed as excessively defensive based on the validity scales were given an opportunity to retest. During the second test administration, participants received altered instructions that explained response distortion. Following the readministration, 83% of participants that received the altered instructions obtained valid profiles on the test.

Item Formats

Edwards (1957) stated that the format of a test item is at least part of the cause of socially desirable responding. The two dominant forms of response format are dichotomous and multichotomous formats (DeVellis, 2003). Dichotomous response formats usually use the forced-choice (or paired-comparison) format and most multichotomous formats use Likert-type or semantic differential scales with three or more scale points (DeVellis). Both response formats have advantages and disadvantages.

Paired-Comparison. A paired-comparison (or forced-choice) scale requires the respondent to choose between a pair of adjectives (i.e., shy or talkative) or behavioral statements. Two examples of paired-comparison scales are given below.

SHY ☐ ☐ **TALKATIVE**

I tend to keep to myself in social situations	☐	☐	**I enjoy talking to others at parties.**

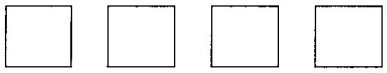
In the case of personality testing, a respondent selects the adjective from the pair that best describes himself or herself. The viability of forced-choice items for personality measurement was documented by Gordon (1951) in an early validity study. As previously discussed, the forced-choice technique has been proposed as a means for controlling socially desirable responding. However, Dunnette et al. (1962) presented research indicating that forced-choice, self-descriptive checklists can be faked. Despite these findings, some researchers remain optimistic that such response formats can reduce socially desirable responding. Jackson, Wroblewski, and Ashton (2000) explained that forced-choice questionnaires can minimize the problem of faking in employment testing

by making the task of responding in a desirable fashion more difficult. In addition, respondents who are motivated to make a positive impression can be forced to choose between items similar in perceived relevance to the job, resulting in a reduction in impression management. Although research has examined forced-choice items and their ability to reduce intentional response distortions (i.e., impression management), few studies have addressed the impact of forced-choice items on unintentional response distortion (i.e., self-deception).

Paired-comparison and dichotomous response formats in general have a number of advantages and disadvantages. An advantage to such items is that they usually do not take much time for a respondent to answer, allowing the completion of more items in a shorter time period (DeVellis, 2003). One disadvantage of paired-comparison items is that they typically have unbalanced response distributions, reducing variability. The end result is a larger number of required items as compared to other item formats (DeVellis). Another disadvantage is that sometimes respondents who are forced to choose between two adjectives equal in social desirability will show some degree of frustration (Edwards, 1957). The resistance that may follow such a choice was examined by Waters and Wherry (1961). They found that subjects were more favorable toward a response format that allowed them to indicate the degree of applicability of each statement in the forced-choice pairs, despite being forced to choose one statement as being more applicable.

Semantic Differential. Semantic differential items use adjective based or behavioral statement endpoints that are bipolar in nature (Osgood, Suci, & Tanenbaum, 1957). The scales are of a multichotomous nature and usually use between five and nine

scale points (DeVellis, 2003). An example of an adjective based semantic differential item is given below.



SHY **TALKATIVE.**

An advantage of the semantic differential and multichotomous scales in general is that they create more scale variance relative to a dichotomous scale with fewer items (DeVellis, 2003), offering economy in terms of both subject and experimenter effort at the data gathering and analysis stages of research (Everett, 1973). Research is mixed on the validity of the semantic differential for personality research. Although some research indicates the semantic differential format is inappropriate for personality assessment due to costs in accuracy of measurement (Everett), others support the technique as a reliable and valid measure if properly constructed as regards wording of scale points, format, and number of scale points (DeVellis).

Comparing Bipolar Adjective Scaling

Paired-comparison and semantic differential scaling are very similar techniques. Both formats use bipolar adjective pairs or behavioral statements with the primary difference being the number of points between the adjectives. Oetting (1967) noted that the semantic differential is essentially a forced-choice technique since the respondent is required to choose between the two adjectives. The difference between the techniques is that the semantic differential allows the respondent to partially endorse an item whereas the paired-comparison requires full endorsement of an adjective. Oetting further indicated

that almost any concept could be rated with very few blanks as long as appropriate adjective pairs were selected.

Research has supported the validity of bipolar adjective scaling techniques for personality assessment. King, King, and Klockars (1983) reported that rating scales formed using bipolar adjectives appear to have potential for measuring traits in a valid manner with very few items. They cautioned, however, that some traits are more amenable to this form of measurement than others. In particular, traits for which the respondent is actually the object of an action by others may be too complex to portray with a simple stimulus such as a pair of bipolar adjectives. For example, in work groups where members perform as a team, assessing personality traits of an individual may be too complex because their work depends on the work of others. Other research indicated that bipolar adjective scales are more reliable and valid when both adjectives are desirable because the influence of social desirability would be less pronounced (Klockars, 1979). Additionally, respondents most frequently chose this method as representing the best type of scale (Klockars).

A number of studies have examined differences among various bipolar adjective scaling techniques. Research has primarily focused on the number of points between the adjective pairs. Bartlett, Quay, and Wrightsman (1960) compared attitude scale items based on a 3-point Likert-type scale with the same items based on a forced-choice scale. The items were administered to two different groups of employees at state mental institutions to measure attitude changes toward mentally retarded persons following training. Bartlett et al. concluded that the correlation between the two measurement techniques indicated two findings. First, the scales were measuring attitude changes

differently. Second, the scales shared some variance. The variance in common to the two techniques likely represents the measurement of attitude in question whereas the unshared variance may occur due to biases.

Comrey and Montag (1982) compared the factor analytic results of two-choice and seven-choice personality item formats. Results indicated higher factor loadings for the seven-choice format, suggesting its superiority over the two-choice format. King et al. (1983) compared the factor structures of dichotomous and multipoint (7-point) bipolar scales. Results indicated that the factor structures between the two methods were very similar with multipoint scales showing a slightly more defined structure than dichotomous scales. The results also suggested that bipolar adjective scaling can be a highly internally consistent measure free from social desirability even with a relatively small number of items. Klockars, King, and King (1981) also suggested that it is possible to obtain internally consistent measures of personality with a small number of bipolar adjective items when the content dimensions are well-defined.

The Present Study

Previous research on social desirability and response formats has focused primarily on reducing socially desirable responding of an intentional nature (i.e., faking good). Little if any research has focused on the forms of socially desirable responding for responses that are not deliberate (i.e., self-deceptive positivity). Self-deceptive positivity is a self-denial or a lack of awareness of negative characteristics about the self. As a result, individuals who respond in this way tend to present themselves in a positive manner. The response is not an intentional deception towards others and is likely not even noticed by the respondent. In the present study, self-deceptive positivity is difficult

to identify with confidence. However, if the groups have no reason to distort their responses intentionally, a difference in responses based on item format would likely indicate such responding. The present study examines the effects that item format (paired comparison versus semantic differential) has on the willingness of respondents to report socially undesirable, self-descriptive information.

I hypothesized that respondents completing the semantic differential version of the instrument will report more socially undesirable information than will respondents who complete the paired-comparison version. This hypothesis is based on the notion that when using the semantic differential technique, respondents are not required to fully endorse the socially undesirable adjective, whereas the paired-comparison requires full endorsement. This reasoning follows from research by Waters and Wherry (1961) who reported that respondents were more favorable toward response formats where the degree of applicability could be indicated. The semantic differential technique allows respondents to indicate some degree of applicability by allowing them to partially endorse the adjective. Subsequent to the analysis of the first sample of data (the police sample), but before I collected data from the university sample, I developed a second hypothesis which states that a semantic differential form of test administration will offer better internal consistency than will a paired-comparison format.

Method

Participants

There were two samples of participants in this study. The first sample consisted of 90 police officers who were participating in promotional testing for Police Sergeant in a large metropolitan city. The full group of officers from which the sample was derived consisted of 135 officers (93 males, 18 females, 23 unreported) with a mean age of 35.57 years (SD = 16.92) and mean length of service of 8.19 years (SD = 1.74). Demographic data were not available for the actual sample of officers who completed the questionnaire. All participating officers had passed the written exam and were in the second stage of the promotional process, job simulations. The second sample consisted of 152 introductory psychology students (79 males and 73 females) with a mean age of 19.49 years (SD = 2.23) from a medium-sized Southeastern university. Students received extra credit for their participation.

Materials

A unidimensional personality assessment of conscientiousness was used. The measure was developed from the conscientiousness facet scale adjective list from the revised version of the NEO-PI (Costa & McCrae, 1992). Two versions of the scale were developed. One version placed the paired bipolar adjectives in a paired-comparison format in which participants must choose one word from the adjective pairs that best describes them. The other version used a semantic differential format in which participants choose one of four options between the bipolar adjective pairs to describe themselves. In order to make results from both forms comparable, a neutral response was not available as an option. Both versions of the instrument consisted of 21 items and

contained the same bipolar adjective pairs in the same order. The instruments can be found in Appendix A (paired-comparison) and Appendix B (4-point semantic differential).

Procedure

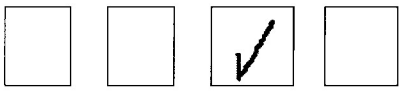
Participants in the police sample were in their second stage of promotional testing. All participants had received a passing score on the written exam and had completed the first day of the oral exam. On the second day of oral exams, participants received a consent form and a version of the instrument (randomly assigned) to complete while they were waiting to enter the testing room. The police officers were informed that all response were anonymous. As such, their responses would not and could not be used to make promotion decisions. Upon completion of the consent form and instrument, participants received a short written summary explaining the purpose of the research and were given the opportunity to ask questions.

Participants in the university sample completed the survey during their class session. Participants received a consent form and were requested to read and sign the form. Participants were informed that all responses would remain anonymous. Upon completion of the consent form, both versions of the instrument were randomly distributed. After completion, participants returned survey materials. Participants were then given a short written summary explaining the purpose of the research and offered the chance to ask questions about the study.

Analysis

Responses on the semantic differential form of the conscientiousness scale were scored in a dichotomous fashion, "0" if they agreed with the response indicating a non-

conscientious behavior (even in part) and "1" if they agreed with the conscientious behavior (again, even in part). I dichotomized in order to compare the scores with the paired-comparison form of the questionnaire which is dichotomous in its existing form. Below, an example of an item is provided.



ORGANIZED **UNORGANIZED**

For this item, the response would be coded as a "0" because the response is on the side of the scale that relates to the adjective *Unorganized,* which is in the non-conscientious direction. Both hypotheses were tested twice, once for the university sample and a second time for the police sample. For the first hypothesis, data were analyzed with an independent samples *t*-test (one tailed, in accordance with the hypothesis). For the second hypothesis, the data were analyzed using the *z* test for differences between correlations from independent samples.

Results

At least one item was left unanswered on 2 of the 92 questionnaires in the police sample and 1 of the 152 questionnaires in the university sample. Cases with missing data were deleted, resulting in 90 complete responses in the police sample and 151 complete responses in the university sample. Total scores were formed by simply summing the point totals across all items. The resulting composite score indicated the number of socially desirable responses that were endorsed. Table 1 summarizes the group means and standard deviations for the two samples.

Table 1

Group Means and Standard Deviations by Sample and Item Format

	Item Format					
	Paired-Comparison			Semantic Differential (4-point)		
Sample	*M*	*SD*	*N*	*M*	*SD*	*N*
Police	18.39	2.19	44	18.00	2.52	46
University	13.72	3.60	76	13.61	4.08	75

Note. Items 4, 5, and 15 removed from the analysis of university sample.

I hypothesized that test items presented in a semantic differential format would yield greater endorsement of socially undesirable behaviors than would the same items presented in a paired-comparison format. To evaluate the hypothesis, I conducted an independent samples *t*-test for each of the sample types to compare the mean differences between the two item formats. For the police sample, analyses indicated no differences

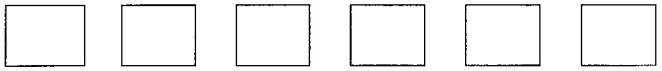
between the paired-comparison and 4-point semantic differential versions of the instrument, $t(88) = .775$, $p > .05$. Examination of the mean scores revealed that participants who completed the paired-comparison format (M = 18.39) differed little in the number of socially desirable responses endorsed as compared to participants who completed the 4-point semantic differential format (M = 18.00).

Because the police data were collected before the university data, I had the opportunity to revise the instrument before its administration to the university sample. Following the previously reported analysis of the police sample, I conducted an internal consistency item analysis. Three items (Items 4, 5, and 15) were eliminated because they did not appear to be measuring the same construct (i.e., low item-total correlations). These three items were eliminated for all subsequent analyses. During this item analysis, I noticed that the semantic differential (dichotomously scored) version had a coefficient alpha that was nominally, though not significantly, stronger than the paired-comparison form of the instrument (.67 versus .60).

Analysis of the university sample yielded results similar to the police sample. Differences between the response formats were nonsignificant, $t(149) = .176$, $p > .05$. Examination of the mean scores revealed that participants who completed the paired-comparison format (M = 13.72) differed little from participants who completed the 4-point semantic differential format (M = 13.61). In summary, for both the police and university samples, results indicated that there were no differences between paired-comparison and 4-point semantic differential item formats in terms of the number of socially desirable items endorsed. Thus, the first hypothesis was not supported.

To test the second hypothesis, I computed coefficient alpha estimates of reliability for each test type within the university sample. Coefficient alphas were .86 for the 4-point semantic differential (dichotomously scored) format and .81 for the paired-comparison format. Unfortunately, the difference between correlation coefficients was not significant, $z = 1.0, p > .05$. Thus, the second hypothesis was not supported.

Following these findings, I considered the possibility that the difference between a 2-point and a 4-point scale was not large enough for respondents to cross over the midpoint of the scale separating a socially desirable response from a socially undesirable response. It is possible that I might find support for the first hypothesis by comparing the paired-comparison format to a semantic differential format with more than four anchors. To make that determination, I conducted a follow-up study to examine differences between the number of socially desirable responses for paired-comparison and 6-point semantic differential scales. An example of a 6-point semantic differential scale is shown below.



SHY **TALKATIVE**

Aside from the number of response alternatives on the semantic differential form, all procedures were identical to those used for prior the university sample. This new university sample consisted of 63 Introductory Psychology students (35 males and 28 females) with a mean age of 19.25 years ($SD = 2.05$).

The analyses of the new university sample data indicated that differences between the response formats were nonsignificant, $t(61) = .267, p > .05$. Examination of the mean

scores revealed that participants who completed the paired-comparison format ($N = 31$, $M = 14.65$, $SD = 2.39$) differed little from participants who completed the 6-point semantic differential format ($N = 32$, $M = 14.44$, $SD = 3.64$).

As with the previous university sample, coefficient alpha estimates of internal consistency reliability were computed to determine whether the 6-point semantic differential form of the questionnaire offered greater internal consistency than the paired-comparison form of the questionnaire. Coefficient alpha estimates of internal consistency were .59 for the paired-comparison version and .83 for the 6-point semantic differential (dichotomously scored) version of the test. The difference between these correlations was significant in the hypothesized direction, $z = 1.96$, $p < .05$. Thus, use of a 6-point semantic differential item format with a dichotomous scoring scheme offers greater internal consistency than a paired-comparison format version of the same set of questions.

Discussion

The present study investigated the effects that item formats have on test-taker responses. The purpose of the study was to see if offering respondents the opportunity to partially endorse an adjective would change their responses. The results of the study indicate that the choice of scaling format makes little difference regarding socially desirable responding. Individuals tend to respond similarly whether the scale is in a paired-comparison format or a semantic differential format.

Different results were found as regards the second hypothesis. The results of the present study indicate that the choice of scaling format does impact the internal consistency of the scale. A bipolar adjective scale using a 6-point semantic differential format was more internally consistent than a scale using the paired-comparison format. Thus, scales that allow more variation in response options tend to provide better quality data even when scored dichotomously. As such, it appears the semantic differential scale is the better choice of the two scales.

Limitations

One limitation of the present study relates to the lack of a midpoint on the semantic differential scale. To compare the semantic differential format with the paired-comparison format, it was necessary to dichotomize the semantic differential version of the scale. Consequently, a neutral response option was not offered. According to the research literature, most semantic differential scales tend to have an odd number of anchors so that the respondent can select a midpoint response, indicating neutrality. As a result, the findings from the present study may not generalize to applications in which a neutral response is offered.

Future Research

Similar to previous research on bipolar adjective scaling (King et al., 1983; Klockars et al., 1981), the results of this study demonstrate the high internal consistency of the bipolar adjective scaling method for personality assessment. As the results of the follow-up study indicated, it appears possible to be able to increase the internal consistency of a bipolar scale simply by using a 6-point semantic differential rating form with the items scored in a dichotomous fashion. Future research should further investigate the degree to which adding scale points will increase the internal consistency of a scale. As it is likely that at some point adding additional points on the scale will cease to provide an increase in internal consistency, future research should determine the optimal number of scale points.

The present study examined response formats in relation to a unidimensional scale of conscientiousness. Future studies should address differences in response formats for other personality dimensions. Furthermore, research should examine the impact that response formats have on scales that assess multiple personality traits. Personality is a complex domain that contains innumerable constructs. Because an individual's personality cannot be summed up in one trait, the interaction of traits should be examined.

Another area in which further research may be fruitful relates to the type of anchor used for the endpoints. The present research examined the properties of scales where adjectives were used as endpoints. Future research should examine scales using behavioral statements (e.g., "I keep my belongings clean and neat") as endpoints. Finally, in relation to bipolar adjective scaling, future research should consider the use of

adjective pairs that are more balanced in terms of social desirability. Most of the adjective pairs in the present study consisted of adjective pairs where one adjective of the pair was much more desirable in a way that picking the undesirable member of the pair was quite unlikely. For example in the adjective pairing, *Mature - Immature*, respondents were very unlikely to endorse the *Immature* adjective because being immature is likely to be considered socially undesirable to most individuals. With such an imbalance in the degree of social desirability between the adjectives, it is unlikely that many respondents would ever consider themselves to be immature. Finally, given the previously mentioned limitation regarding the lack of a midpoint of the semantic differential scale, future studies should examine differences in socially desirable responding when a scale is provided with a midpoint anchor. Doing so, however, will require a scoring system different from the one used in this study.

Summary

Although personality testing is a popular and widely-used method for personnel selection, researchers have criticized its use given the potential for response distortions by applicants (Ones et al., 1996). Researchers have developed many techniques to control for such response distortions including the use of different response formats. Using both a university sample of introductory psychology students and an applied sample of police officers during a promotional exam, the present study examined two bipolar adjective scaling methods (paired-comparison and semantic differential) to determine format effects on respondents' tendency to report socially undesirable information on a self-report personality measure of conscientiousness. Results indicate that no differences exist in socially desirable responding based on the scaling method used to assess personality.

Furthermore, a follow-up study using college students and a semantic differential scale with six points also failed to find a difference. Results did indicate differences in internal consistency between a 6-point semantic differential scale and a paired-comparison scale. Although these results appear to suggest a clear solution for scale format, this research addresses only one area for improving scale development. Consequently, future research is needed in order to address the many complex issues inherent in personality assessment.

References

Anastasi, A. (1982). *Psychological Testing* (5th Ed.). New York: MacMillan Publishing Company.

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261-272.

Bartlett, C. J., Quay, L. C., & Wrightsman, L. S., Jr. (1960). A comparison of two methods of attitude measurement: Likert-type and forced-choice. *Educational and Psychological Measurement, 20*, 699-704.

Butcher, J. N. (1994). Psychological assessment of airline pilot applicants with the MMPI-2. *Journal of Personality Assessment, 62*, 31-44.

Cigrang, J. A., & Stall, M. A. (2000). Readministration of the MMPI-2 following defensive invalidation in a military job applicant sample. *Journal of Personality Assessment, 76*, 472-481.

Comrey, A. L., & Montag, I. (1982). Comparison of factor analytic results with two-choice and seven-choice personality item formats. *Applied Psychological Measurement, 6,* 285-289.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO-PI-R professional manual*. Odessa, FL: Psychological Assessment Resources.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description checklist. *Personnel Psychology*, *15*, 13-24.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: The Dryden Press

Everett, A. V. (1973). Personality assessment at the individual level using the semantic differential. *Educational and Psychological Measurement*, *33*, 837-844.

Gordon, L. V. (1951). Validities of the forced-choice and questionnaire methods of personality measurement. *Journal of Applied Psychology*, *35*, 407-412.

Jackson, D.N., Wroblewski, V.R., & Ashton, M.C. (2000). The impact of faking on employment tests: Does forced-choice offer a solution? *Human Performance, 13*, 371-388.

Klockars, A. J. (1979). Evaluative confounding in the choice of bipolar scales. *Psychological Reports*, *45*, 771-775.

Klockars, A. J., King, D. W., & King, L. A. (1981). The dimensionality of bipolar scales in self-description. *Applied Psychological Measurement*, *5*, 219-227.

King, D. W., King, L. A., & Klockars, A. J. (1983). Bipolar adjective scales for self-description: Reliability and validity data. *Educational and Psychological Measurement*, *43*, 879-886.

King, L. A, King, D. W., & Klockars, A. J. (1983). Dichotomous and multipoint scales using bipolar adjectives. *Applied Psychological Measurement*, *7*, 173-180.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*, 263-280.

Nicholson, R. A., & Hogan, R. (1990). The meaning of personality test scores. *American Psychologist, 43*, 621-626.

Oetting, E. R. (1967). The effect of forcing response on the semantic differential. *Educational and Psychological Measurement*, *27*, 699-702.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660-679.

Osgood, C. E., Suci, G. J., & Tanenbaum, P. H. (1969). *The measurement of meaning*. In Snider, J. G. & Osgood, C. E. (Eds.). *Semantic differential technique: A Sourcebook*. Chicago: Aldine Publishing Company.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609.

Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin*, *21*, 100-108.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, *60*, 307-317.

Robins, R. W., & Paulhus, D. L. (2001). The character of self-enhancers: Implications for organizations. In Roberts, B.W. and Hogan, R. (Eds.), *Personality psychology in the workplace* (pp. 193-219).Washington, D.C.: American Psychological Association.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*, 634-644.

Strong, D. R., Greene, R. L., & Kordinak, S. T. (2002). Taxometric analysis of impression management and self-deception in college student and personnel evaluation settings. *Journal of Personality Assessment*, *78*, 161-175.

Waters, L. K., & Wherry, R. J. (1961). Evaluation of two forced-choice response formats. *Personnel Psychology*, *14*, 285-289.

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review*, *12*, 250-264.

Appendix A:

Paired Comparison Scale

DIRECTIONS: Below are 21 adjective pairs that can be used to describe a person's personality. For each adjective pair, please mark the box next to the adjective that best describes you.

EXAMPLE 1:

TALKATIVE	☐	☑	**SHY**

This person marked the "SHY" box because he felt the adjective "SHY" described him better than the adjective "TALKATIVE".

EXAMPLE 2:

OUTSPOKEN	☑	☐	**RESERVED**

This person marked the "OUTSPOKEN" box because she felt the adjective "OUTSPOKEN" described her better than the adjective "RESERVED".

1	**AMBITIOUS**	☐	☐	**SATISFIED**
2	**INSECURE**	☐	☐	**CONFIDENT**
3	**ORGANIZED**	☐	☐	**UNORGANIZED**
4	**IMPULSIVE**	☐	☐	**DELIBERATE**
5	**PATIENT**	☐	☐	**IMPATIENT**
6	**HASTY**	☐	☐	**CAREFUL**
7	**MATURE**	☐	☐	**IMMATURE**
8	**UNMOTIVATED**	☐	☐	**INDUSTRIOUS**

9	**DETERMINED**	☐	☐	**WAVERING**
10	**PASSIVE**	☐	☐	**ENERGETIC**
11	**FOCUSED**	☐	☐	**DISTRACTED**
12	**INEFFICIENT**	☐	☐	**EFFICIENT**
13	**CARELESS**	☐	☐	**CAREFUL**
14	**PRECISE**	☐	☐	**VAGUE**
15	**DEFENSIVE**	☐	☐	**OPEN**
16	**HARDWORKING**	☐	☐	**LAZY**
17	**UNCREATIVE**	☐	☐	**RESOURCEFUL**
18	**ATTENTIVE**	☐	☐	**FORGETFUL**
19	**INACTIVE**	☐	☐	**ACTIVE**
20	**DISCIPLINED**	☐	☐	**UNDISCIPLINED**
21	**INCONSISTENT**	☐	☐	**PERSISTENT**

Before returning this questionnaire to the administrator, please make sure you have responded to all 21 adjective pairs.

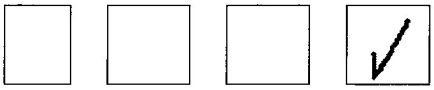
Thank you for your participation.

Appendix B:

Semantic Differential (4-Point) Scale

DIRECTIONS: Below are 21 adjective pairs that can be used to describe a person's personality. For each adjective pair, please mark the box that best describes where you would fall on the continuum between the adjective descriptions.

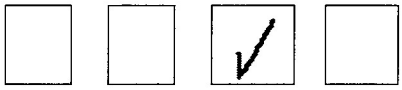
EXAMPLE 1:



APATHETIC **ENTHUSIASTIC**

This person marked this box because he felt the adjective "ENTHUSIASTIC" described him better than the adjective "APATHETIC" because he feels he is always interested and has feelings about different issues.

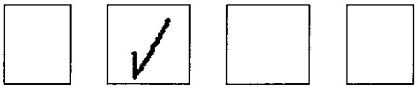
EXAMPLE 2:



TALKATIVE **SHY**

This person marked this box because he felt the adjective "SHY" described him better than the adjective "TALKATIVE", however, he did not feel he was completely "SHY".

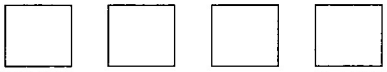
EXAMPLE 3:



OUTSPOKEN **RESERVED**

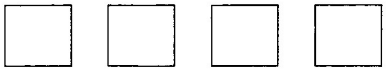
This person marked this box because she felt the adjective "OUTSPOKEN" described her better than the adjective "RESERVED", however, she did not feel she was fully "OUTSPOKEN".

1



AMBITIOUS **SATISFIED**

2



INSECURE **CONFIDENT**

3	☐	☐	☐	☐	
	ORGANIZED			**UNORGANIZED**	
4	☐	☐	☐	☐	
	IMPULSIVE			**DELIBERATE**	
5	☐	☐	☐	☐	
	PATIENT			**IMPATIENT**	
6	☐	☐	☐	☐	
	HASTY			**CAREFUL**	
7	☐	☐	☐	☐	
	MATURE			**IMMATURE**	
8	☐	☐	☐	☐	
	UNMOTIVATED			**INDUSTRIOUS**	
9	☐	☐	☐	☐	
	DETERMINED			**WAVERING**	
10	☐	☐	☐	☐	
	PASSIVE			**ENERGETIC**	
11	☐	☐	☐	☐	
	FOCUSED			**DISTRACTED**	
12	☐	☐	☐	☐	
	INEFFICIENT			**EFFICIENT**	
13	☐	☐	☐	☐	
	CARELESS			**CAREFUL**	
14	☐	☐	☐	☐	
	PRECISE			**VAGUE**	

15 ☐ ☐ ☐ ☐
DEFENSIVE **OPEN**

16 ☐ ☐ ☐ ☐
HARDWORKING **LAZY**

17 ☐ ☐ ☐ ☐
UNCREATIVE **RESOURCEFUL**

18 ☐ ☐ ☐ ☐
ATTENTIVE **FORGETFUL**

19 ☐ ☐ ☐ ☐
INACTIVE **ACTIVE**

20 ☐ ☐ ☐ ☐
DISCIPLINED **UNDISCIPLINED**

21 ☐ ☐ ☐ ☐
INCONSISTENT **PERSISTENT**

Before returning this questionnaire to the administrator, please make sure you have responded to all 21 adjective pairs.

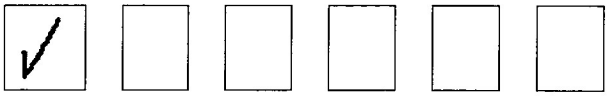
Thank you for your participation.

Appendix C:

Semantic Differential (6-Point) Scale

DIRECTIONS: Below are 21 adjective pairs that can be used to describe a person's personality. For each adjective pair, please mark the box that best describes where you would fall on the continuum between the adjective descriptions.

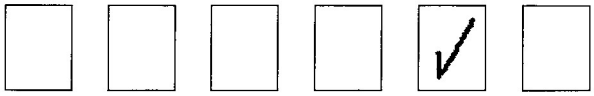
EXAMPLE 1:



ENTHUSIASTIC **APATHETIC**

This person marked this box because he felt the adjective "ENTHUSIASTIC" described him better than the adjective "APATHETIC" because he feels he is always interested and has feelings about different issues.

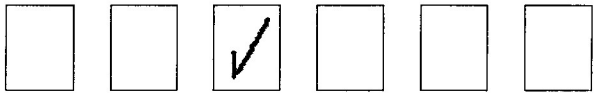
EXAMPLE 2:



TALKATIVE **SHY**

This person marked this box because he felt the adjective "SHY" described him better than the adjective "TALKATIVE", however, he did not feel he was completely "SHY"

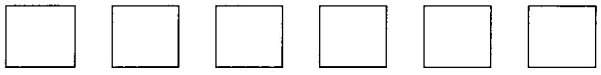
EXAMPLE 3:



OUTSPOKEN **RESERVED**

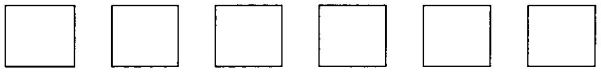
This person marked this box because she felt the adjective "OUTSPOKEN" described her better than the adjective "RESERVED", however, she did not feel she was fully "OUTSPOKEN".

1



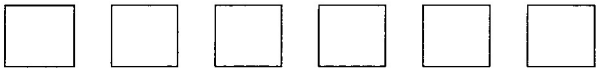
AMBITIOUS **SATISFIED**

2



INSECURE **CONFIDENT**

3



ORGANIZED **UNORGANIZED**

4 ☐ ☐ ☐ ☐ ☐ ☐

IMPULSIVE **DELIBERATE**

5 ☐ ☐ ☐ ☐ ☐ ☐

PATIENT **IMPATIENT**

6 ☐ ☐ ☐ ☐ ☐ ☐

HASTY **CAREFUL**

7 ☐ ☐ ☐ ☐ ☐ ☐

MATURE **IMMATURE**

8 ☐ ☐ ☐ ☐ ☐ ☐

UNMOTIVATED **INDUSTRIOUS**

9 ☐ ☐ ☐ ☐ ☐ ☐

DETERMINED **WAVERING**

10 ☐ ☐ ☐ ☐ ☐ ☐

PASSIVE **ENERGETIC**

11 ☐ ☐ ☐ ☐ ☐ ☐

FOCUSED **DISTRACTED**

12 ☐ ☐ ☐ ☐ ☐ ☐

INEFFICIENT **EFFICIENT**

13 ☐ ☐ ☐ ☐ ☐ ☐

CARELESS **CAREFUL**

14 ☐ ☐ ☐ ☐ ☐ ☐

PRECISE **VAGUE**

15 ☐ ☐ ☐ ☐ ☐ ☐

DEFENSIVE **OPEN**

16 ☐ ☐ ☐ ☐ ☐ ☐

HARDWORKING **LAZY**

17 ☐ ☐ ☐ ☐ ☐ ☐

UNCREATIVE **RESOURCEFUL**

18 ☐ ☐ ☐ ☐ ☐ ☐

ATTENTIVE **FORGETFUL**

19 ☐ ☐ ☐ ☐ ☐ ☐

INACTIVE **ACTIVE**

20 ☐ ☐ ☐ ☐ ☐ ☐

DISCIPLINED **UNDISCIPLINED**

21 ☐ ☐ ☐ ☐ ☐ ☐

INCONSISTENT **PERSISTENT**

Before returning this questionnaire to the administrator, please make sure you have responded to all 21 adjective pairs.

Thank you for your participation.